Report #2
PD-Internship
Lerner Paul
11/03/2019

I mostly read papers about PD and Handwriting (HaW) for now. This report is divided by papers' topic. I highlighted the parts where something is unclear or I doubt…
We also created a GitLab to share the code and the reports.

Machine Learning

I wonder how reliable are the results of Taleb et al. 2017[1] since they select among 1323 features the one that were the most statistically significant (using t-test or Mann-Whitney test) then the one that gave **the best classification performance**. I wonder if their classifier is then overfitted to the data.
Taleb et al. 2018[2] achieved 94%, 92%, and 88% accuracy predicting the H&Y stage, UPDRS scores, and total UPDRS, respectively. This motivates the idea that HaW analysis can be used to monitor PD progression in a cheap, non invasive and possibly "from home" way. Although, once again, I doubt the strength of their results given their small dataset and the process of feature selection and data augmentation.

On March 7th we decided that we will try a pure data-driven approach were we **don't extract any features** from the data, as we're going to use deep learning techniques, unlike most of the SoA. Although we're a priori going to normalize the data before feeding it to the model, like Pereira et al. and Drotar et al. Maybe we could apply a Fourier transform also ?

[1] Feature Selection for an Improved Parkinson's Disease Identification Based on Handwriting Catherine Taleb (1)(2), Laurence Likforman-Sulem (2), Maha Khachab (1), Chafic Mokbel (1)
[2] A Reliable Method to Predict Parkinson's Disease Stage and Progression based on Handwriting and Resampling Approaches. Catherine Taleb, Maha Khachab, and Chafic Mokbel Laurence Likforman-Sulem

Pereira et al. 2016[3] attained 91% sensitivity and 76% specificity, using convolutional neural networks (CNN) by transforming exam signals into a grayscale image (1 row per milliseconds and 6 columns corresponding to the 6 signal channels). They also had very strange results, e.g. 98% sensitivity but 6% specificity with some architectures. This is probably because their dataset is imbalanced. They trained their model on only 2 tasks for now : spiral and meander (their best results are on the latter). It's interesting that they achieved decent results with the Optimum-Path Forest (OPF, that they use as a baseline) algorithm : 87% sensitivity and 71% specificity. Also, their results with OPF are more consistent/robust : neither sensitivity nor specificity gets worse than 50% depending on the task, image size and dataset split ratio, unlike their different CNN architectures.
The paper came out in 2016 and they claim to be the first to diagnose PD using **deep learning** techniques. They chose not to use **transfer learning** techniques since their images are "domain-specific" (they don't mention the PaHaW database nor any work from Drotar et al. in this paper but they do in their latest work and still chose not to use transfer learning for the same reasons).
In a more recent work[4] they achieved 91% sensitivity and 89% specificity by fine-tuning the hyperparameters : learning rate η, penalty parameter (momentum) α and weight decay λ using the BA, PSO and FA techniques. Their best result were achieved with BA. This time, their best results are with the spiral task, and not the meander. Their better results might be explained because of their use of an updated version of the **HandPD** dataset (let's call it **HandPD2**) which is more balanced than the old one.
In their latest work[5] they achieved 97% sensitivity and 94% specificity by taking into account more tasks (i.e. they train a CNN over each task) and predict using **majority voting** to obtain the final result. They say that the CNN parameters were set empirically and strangely don't mention their previous work on CNN finetuning (one can assume that they used the parameters that they found best in this previous work). As a baseline, they also used OPF, SVM and Naive Bayes classifier using the same voting technique but transforming the data with gray level co-occurrence matrices (GLCM)[6]. They also achieved great results with those : 99% sensitivity and 91% specificity. Once again, their better results might be explained because of their use of an updated version of the HandPD dataset : **NewHandPD** which is more balanced than **HandPD** and contains more observations than **HandPD2**.

[3] Deep Learning-aided Parkinson's Disease Diagnosis from Handwritten Dynamics.Clayton R. Pereira, Silke A. T. Weber, Christian Hook, Gustavo H. Rosa, João Papa
[4] Convolutional Neural Networks Applied for Parkinson's Disease Identification
Clayton R. Pereira 1 , Danillo R. Pereira 2 , Joao P. Papa 2(B) , Gustavo H. Rosa 2 , and Xin-She Yang 3
[5] Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. Clayton R. Pereira a , Danilo R. Pereira b , Gustavo H. Rosa c , Victor H.C. Albuquerque d , Silke A.T. Weber e , Christian Hook f , João P. Papa
[6] Haralick RM, Shanmugam K, et al. Textural features for image classification.
IEEE Trans. Syst. Man Cybern 1973;6:610–21.

Passos et al.[7] used a CNN : ResNet-50 (trained for object detection on ImageNet) to extract features from **the images of the HandPD** database which were then fed to a classifier (notice that they didn't use **NewHandPD** even though the paper came out in 2018). The authors experimented with OPF, SVM and Bayes classifiers. A lot of their work is unclear to me :

- why is the "patient accuracy" different from the sensitivity ?? In the same way, why is the "control accuracy" different from the specificity ??

I'm very surprised that they obtain a 92% accuracy using only Res-Net (i.e. by adding just one classification layer after the feature extraction) because the data and the classification task are very different from ImageNet (Pereira et al. chose not to use **transfer learning** because of that). They achieved 97% accuracy with OPF after applying a PCA to reduce the extracted features to a dimensional space of size 100. Their results are more robust than the ones of Pereira et al. 2016 : they deal better with the imbalanced data. It's also impressing that they used only the **images of the HandPD** dataset, and not the sensors data, Pereira et al. 2015[8] had had very poor results doing that.

Rosenblum et al.[9] achieved 95% sensitivity and 100% specificity using discriminant analysis (can we compare those results and the rest of SoA ? e.g. no train-test set, no cross valid). Drotar et al. 2014, 2015 cite this paper only as a *"preliminary data [study that] suggest that handwriting might serve as a diagnostic marker for PD"* even though they end up with worse results than them. Using only 9 features : mean time taken to write each stroke, both on-paper and in-air ; stroke time, both on-paper and in-air ; mean velocity for the entire task in seconds ; mean stroke's length, width, and height in centimeters ; and the mean pressure.
The 40 subjects did only two tasks : write their name and copy an address (same address for all). Patients performed study while on medication.

Handwriting features for PD diagnosis

Stanley et al.[10] and Taleb et al. 2017 averaged the training data of each segment of the same task for each subject. On March 7th we agreed upon keeping each data separately in order to augment the dataset and not lose any information.

---

[7] Parkinson Disease Identification using Residual Networks and Optimum-Path Forest. Leandro A. Passos ∗ , Clayton R. Pereira † , Edmar R. S. Rezende ‡ , Tiago J. Carvalho § ,Silke A. T. Weber ¶ , Christian Hook , João P. Papa

[8] A Step Towards the Automated Diagnosis of Parkinson's Disease: Analyzing Handwriting Movements, Clayton R. Pereira, Danillo R. Pereira, Francisco A. da Silva, Christian Hoo

[9] Handwriting as an objective tool for Parkinson's disease diagnosis
Sara Rosenblum • Margalit Samuel • Sharon Zlotnik • Ilana Erikh • Ilana Schlesinger 2013

[10] Digitized Spiral Analysis is a Promising Early Motor Marker for Parkinson Disease. Kaili Stanley 1, Johann Hagenah 3, Norbert Brüggemann 3, Kathrin Reetz 3, Lawrence Severt 1, Christine Klein 3, Qiping Yu, Carol Derby 2,**, Seth Pullman 4, and Rachel Saunders-Pullman 1,2,*,

Drotar et al., 2014[11] found that kinematic features from in-air movements were more discriminative than features from on-surface movements (87% VS 78% sensitivity with a SVM classifier). This is explained by Rosenblum et al. as *"in-air time is a manifestation of* 'planning the next movement'*, as required in the sequential process of handwriting"*.
Drotar et al., 2016[12] also found that the Archimedean spiral task was the least discriminative task although Stanley et. al suggests that it might be an early marker for PD. Drotar et al. explain that it may be because they focused on handwriting structures and didn't use any spiral-specific structures. It would be interesting to see if we get the same results using a neural network with a pure data-driven approach (i.e. without extracting features), Pereira et. al achieved great results using only spiral (but can't compare with handwriting tasks).

Drotar et al., 2015[13] found that their second task was more *"more contributing to overall classification performance"* and they explain that because : *"The second task is a pseudoword not existing in Czech language, whereas tasks from 3 to 6 include words used in everyday language. Therefore, the performance of the second task* ***is not automatized****. The lateral premotor cortex is* ***additionally activated*** *when attention must be paid during a motor task."*. I don't understand, this goes against everything else I've been reading, if the task is not automatized it should be **the less discriminative**. Letanneux et al. recommend *"Asking participants to write a long text should be* ***avoided*** *because it implies either copying the text or writing under dictation, and in both cases,* ***cognitive processes are required****. Participants should write single* ***familiar words****, without any spelling difficulty and syntactic and semantic contexts."*.

Micrographia

Inzelberg et al.[14] report that micrographia is not exclusive to PD and has often been observed by different researchers (cf. article for references) following focal basal ganglia lesions without any accompanying parkinsonism. They made this figure which resumes all the disease which cause micrographia :

[11] Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. Peter Drotár a , Jiří Mekyska a , Irena Rektorová b,∗ , Lucia Masarová b ,Zdenek Smékal a , Marcos Faundez-Zanuy c
[12] Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease
[13] Decision Support Framework for Parkinson's Disease Based on Novel Handwriting Markers
Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, Senior Member, IEEE, and Marcos Faundez-Zanuy, Member, IEEE
[14] Inzelberg, R., Plotnik, M., Harpaz, N. K., & Flash, T. (2016). Micrographia, much beyond the writer's hand. *Parkinsonism & related disorders*, *26*, 1-9.

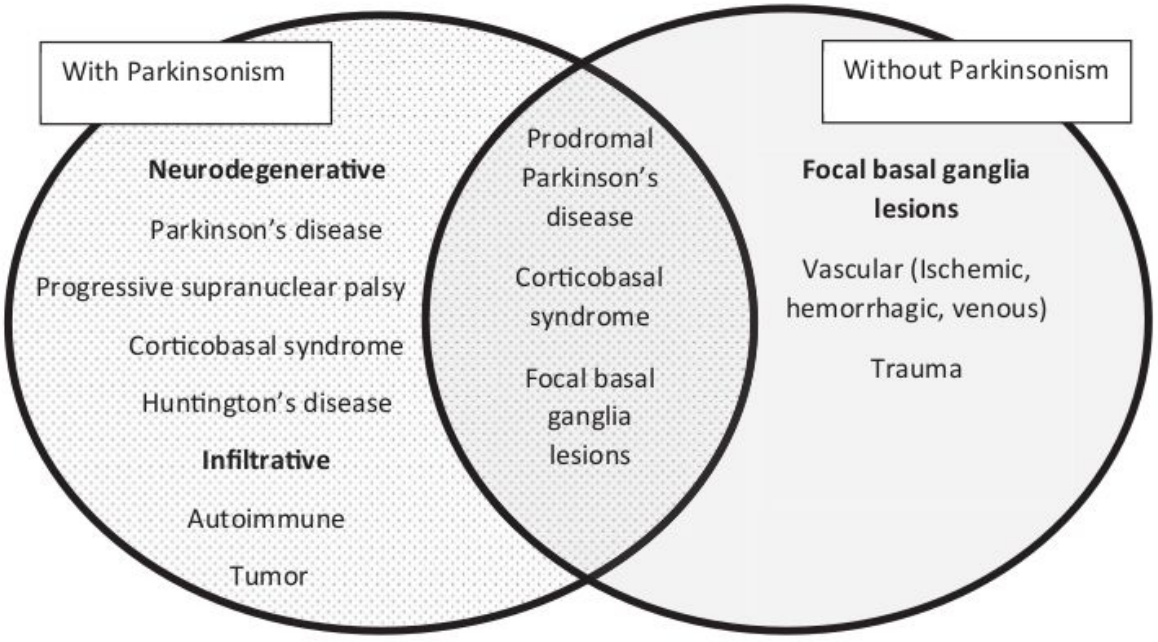

Fig. 2. Etiology of micrographia with and without accompanying Parkinsonism.

McLennan et al.[15] found that micrographia may antedate additional motor signs of PD by three to four years. They also observed that 30% of patients showed micrographia in the course of PD and about 5% reported its occurrence prior to PD diagnosis. Therefore, one should be careful about collecting data from patients who specifically suffer from micrographia (PaHaW and PDMultiMC databases don't mention anything about their subject suffering from micrographia). Although the number of patients showing micrographia varies according to the study's methodology, such as information retrieved through verbal history (about 60% of patients), actual testing (about 50%)[16].
Wu et al. show that asking patients to pay attention to the letter size improves micrographia, therefore, one should be careful to not give any hint to subjects when collecting data. Drotar et al. write that *"subjects were asked to write a sentence at a **self-determined comfortable size** and speed."*.

Dysgraphia

I wonder if dysgraphia has the same properties (cf. Inzelberg et al.) as micrographia :

- 2 types : progressive & consistent

[15] Micrographia in Parkinson's disease J.E.McLennan K.Nakano H.R.Tyler R.S.Schwab
[16] A. Wagle Shukla, S. Ounpraseuth, M.S. Okun, V. Gray, J. Schwankhaus, W.S. Metzer, Micrographia and related deficits in Parkinson's disease: a cross-sectional study, BMJ Open 2 (2012).

- affected by writing directions (i.e. horizontal or vertical)
- antedate additional motor signs of PD
- affected by dominant hand
- affected by language (i.e. native or learned)

Letanneux et al.[17] show that writing velocity and smoothness/fluency abnormalities are more frequent (above 75%) than diminished letter size (between 30%-50%), this is an excellent motivation to analyse kinematic features. They also conclude that *"dopamine depletion would affect not only the amplitude of the motor output, but also its kinematics"*.
Dysgraphia is promising to discriminate between PD and other diseases as Yu et al.[18] reported significant differences in handwriting kinematics between patients with PD and Essential Tremor (ET) ; Ling et al.[19] reported difference between PD and Progressive Supranuclear Palsy (PSP) ; PD patients react to Levodopa unlike non-PD patients so I guess one could collect data from people suffering from different disease both on-drug and off-drug… Although Inzelberg et al. report that L-dopa effects are inconsistent and clinical observations states that only some PD patients restore letter size with dopaminergic therapy (cf. article for references). Also, Wu et al.[20] show that L-dopa improves consistent micrographia but not progressive micrographia.

Datasets

A summary of the different datasets characteristics is available here and in the git repo.

Our Goal - Conclusion

We should state clearly that our goal is to classify healthy and sick people and not people suffering from PD and people suffering from other diseases which cause micrographia (listed in Fig. 2) as we don't have any database containing those (though we should check if it exists and dysgraphia is promising to discriminate between PD and other diseases as explained above). Therefore, our work is an extension of the work of Drotar et al. and Pereira et al. and our goal is to find a better machine learning tool than them. In order to make our work more interesting I think we should really try transfer learning between PaHaW,

[17] A. Letanneux, J. Danna, J.L. Velay, F. Viallet, S. Pinto, From micrographia to Parkinson's disease dysgraphia, Mov. Disord. 29 (2014) 1467e1475.
[18] Yu NY, Van Gemmert AWA, Chang SH. Characterization of grapho-motor functions in individuals with Parkinson's disease and essential tremor. Behav Res Methods 2016;35:795–806.
[19] H. Ling, L.A. Massey, A.J. Lees, P. Brown, B.L. Day, Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease, Brain 135 (2012) 1141e1153.
[20] Wu T, Zhang J, Hallett M, Feng T, Hou Y, Chan P. : Neural correlates underlying Micrographia in Parkinson's disease.

PDMultiMC and NewHandPD. I think I should chat with Catherine and maybe work with her depending on her goals.

Todo List

On March 7th we agreed that it would be interesting that I get into papers :

- which diagnose other diseases based on HaW features
- which diagnose PD using other features than HaW, e.g. speech

In the same direction, I think I should investigate on automatic diagnosis (via HaW ?) for non-PD disease that cause micrographia (listed above).

I also think that I should look into the conferences where the SoA published their papers so maybe we could attend to one.